

MEDICAL EVIDENCE AND CLINICAL PRACTICE: HOW CAN
TECHNOLOGY ASSESSMENT NARROW THE GAP?

Stan N. Finkelstein*
Peter Temin*

October 1982
WP# 1355-82

ABSTRACT

Historically, the link between the evidence that a medical technology is worthwhile and its acceptance or rejection by decision-makers has not been an especially strong one. The current interest in "technology assessment" has aimed to alter that state of affairs and to narrow the gap. Technology assessment seeks to settle a controversy through analysis, evaluation, or consensus. But, settlement of the controversy does not necessarily result in changes in practitioner behavior. We propose an agenda for an assessment of assessments themselves in order to learn how to enhance the relationship between scientific evidence and medical practice and ultimately, to improve the quality of medical care.

MEDICAL EVIDENCE AND CLINICAL PRACTICE: HOW CAN TECHNOLOGY ASSESSMENT NARROW THE GAP?

I. INTRODUCTION

It is commonly stated that remarkable technological advances in medical diagnosis, treatment, rehabilitation, and prevention have taken place since the second World War. Only recently, however, has the extent to which medical or scientific evidence justifies their adoption and use been seriously questioned.

Historically, the relationship between the evidence that a medical technology is worthwhile and its acceptance or rejection by decision-makers has not been an especially strong one. However, the current interest in "technology assessment," that has involved such institutions as the National Institutes of Health, the Institute of Medicine and a number of groups and individuals from academia, industry, and government, has aimed to alter that state of affairs and narrow the gap. By technology assessment, we refer to policy-mandated, systematic attempts to advance the efficacious use of medical technologies and practices.

In this article we explore some of the determinants of clinical practice and examine the manner in which medical evidence and technology assessment change the treatments or diagnostic tests that physicians prescribe for their patients. We consider the mechanism through which four current policies, all forms of technology assessment, could help to make clinical practice more scientific.

II. CLINICAL PRACTICE AND ITS DETERMINANTS

When a patient presents with a set of symptoms and signs which the clinician recognizes as a familiar pattern, the response is to order appropriate diagnostic tests or treatment. Each practitioner probably has a collection of algorithms for dealing with those recognizable patterns. The summation of one individual's algorithms over the universe of patient presentations might be referred to as that physician's clinical practice.¹

Clinical practice by a typical doctor would be expected to evolve over time based upon education, practical experience, and the collective experience of professional colleagues. Algorithms which include the use of drugs, procedures and other technologies are continually updated or discarded according to new information. But how much of the information is based upon medical or scientific evidence remains an open question.

In order for medical evidence to influence the adoption of a new technology or the abandonment of an established one, there needs to be an information base, generated through research or other means, which describes the clinical application of the technology in question. That information needs to be interpreted and then communicated to the practitioners to whom it relates. A physician may then make a tentative decision to accept or reject the information as a basis for changing practice. A tentative change may become a permanent change after "learning by doing" corroborates its value. If these events occur in an effective and timely fashion, then the relationship between medical evidence and clinical practice should be quite strong.

Technology assessment potentially influences acceptance or rejection of clinical practices at two points in the process. First, it can occur after the generation of the research base, but before practitioners can gain access to the technology for use on their patients. Second, there also is a role for technology assessment later on, after a technology has been adopted and may even be in wide use.

III. THE GAP BETWEEN MEDICAL EVIDENCE AND CLINICAL PRACTICE

In reality, however, each of the above described determinants of change in clinical practice is prone to defects which serve to weaken the influence of medical evidence. The information base (assuming that it exists) itself can be defective because of poor research or study design. The information can be of high quality, but the object of misinterpretation. Communication of the information base may be either inadequate or misdirected. And, there may be an excessive reliance on the observations made while "learning by doing." We discuss each of these problems in turn.

A. Defects in the information base or its interpretation.

Inadequacies in clinical information bases have received wide attention in recent years. There are at least two aspects to this problem. First, for many technologies, is the paucity of evaluative data which are based on legitimately designed efficacy trials. The second aspect is the considerable room for misinterpretation of the data generated from even well-designed research studies.

The quality of the clinical research bases has been addressed by others. Fletcher and Fletcher considered the trends in designing clinical research studies over the 30 year period from 1945 - 75 and

confirmed the current emphasis on randomized controlled efficacy trials.² Amounting to 10% of clinical research recently reported in major journals, studies of this type were essentially non-existent thirty years ago. But, disappointingly, they reported far greater reliance today on single case reports than on the usually better designed case control and cohort studies as compared to thirty years ago. Fineberg, in his study of the evolution of the gastric freezing technique to treat ulcer disease, found that efficacy data deriving from decently designed studies did not become available until well after the technique had become widely accepted and then largely rejected.³ The clinical research data that eventually was reported failed to find the technology to be efficacious.⁴ Wennberg and Bunker describe substantial variation in rates of utilization of various common surgical procedures in different populations without significant effects on aggregate health measures of these populations and without apparent information bases with which to establish selection criteria for patients who might undergo those procedures.⁵

There have also been problems noted with the use of information deriving from clinical research that had been properly designed but not adequately interpreted. Eddy describes and catalogues a number of errors in analyzing and interpreting clinical research studies.^{1,6} An example would be the errors in probabilistic reasoning in which he found researchers to have systematically confused the sensitivity of mammography in the diagnosis of breast cancer with its predictive value in discriminating between the presence and absence of disease.

B. Pitfalls in the communication of the information base

Even if the information and research bases are adequate and they have been properly interpreted, they need to be effectively communicated to the target practitioners in order for them to be important determinants of clinical practice. Defects in the communication process can be of at least three possible kinds. The wrong information can be given to the right people; the right information can be targeted to the wrong audience; or, the wrong information can be communicated to the wrong audience.

Researchers have in the past been interested in finding out through what communication channels doctors become informed of the existence of new technologies.^{7,8,9,10} There has been less attention paid to the content of this information and, in particular, how much of it relates to medical evidence.^{11,12}

The classic study of the communication of a pharmaceutical innovation was carried out by Coleman, Katz, and Menzel in the 1950s.⁹ This major empirical study examined the adoption process associated with a new antibiotic drug, "gammanym", the third in a series of related pharmaceuticals, following "alphanym" and "betanym". Their study utilized prescription records of physicians practicing in three mid-western communities, and also included structured interviews with the physicians about the determinants of their drug adoption behavior, the nature of their practices, and the structure of the colleague networks with whom or through whom they communicated. The findings from their study suggested a two-step process by which physicians learn about and go onto adopt new drugs. The first step involves opinion leaders, highly respected physicians practicing in the communities, who find out about

the new drug innovations from external sources, predominantly drug manufacturers' promotional activity. The bulk of the practitioners, however, adopted gammanym as a result of a "chain reaction"; that is, they learned about it from the colleagues with whom they regularly communicated informally.

Gammanym and its predecessor drugs were not identified by name in the Coleman, Katz, and Menzel study. We believe them to have corresponded to tetracycline and two of its chemically similar derivatives. It is recognized today that the pharmacologic actions of and indications for all three agents are quite similar. The communication patterns that were responsible for effecting the adoption of gammanym therefore either were independent of any real information about the relative efficacies of the alternative drugs, or the popular perceptions and interpretations of that information base were wrong.

Another potential problem in the communication of information about medical practices that may lead to change in clinical practice is that the information may be communicated via channels that do not reach a substantial fraction and possibly the most important segment of the target audience. Allen and others have written about the structure of professional communication among scientists as compared to engineers.¹³ Scientists doubtless are concerned about the desire to disseminate the substance of their scientific findings, but tend also to seek to maximize recognition because professional career development depends largely upon it. Engineers, on the other hand, may be relatively more concerned with communications directed toward solutions to technical problems leading to practical applications. Informal types of communication may be more important in achieving the objectives of the engineers.

Clinical researchers very likely see themselves as "scientists" and seek to publish their finding in prestigious journals in order to maximize recognition. These research-oriented journals, as Young points out, may have greater interest in describing research methodology than in communicating the details of guidelines for implementing change in clinical practice.¹² Also, these research-oriented journals may not be the best places to publish if the objective is to reach a specific audience of practitioners because the journals may not be read by important segments of the target audience.

C. Excessive Reliance on "learning by Doing"

Of a national sample of practitioners surveyed by Finkelstein and colleagues, over three-fourths considered their own experiences with the use of drugs and other medical technologies to be a highly important determinant of their clinical practices.¹⁴ "Learning by doing" has its place in certain aspects of medical practice, especially those which require the acquisition of skills that become refined with experience. Luft, Bunker, and Enthoven found an important empirical relationship between experience and outcome for a selection of surgical procedures.¹⁵ However, in many other kinds of practice, there may be a tendency to place too much reliance on one's own personal observations of the use of a drug, procedure, or other technology.

In some sense, medical practitioners appear to be involved in a sort of clinical feedback loop. Conducting ongoing, uncontrolled, unrandomized studies, they rely at least partly on their own experiences and those of their colleagues to update expectations about the clinical value of specific treatments. Provided that the feedback is not

significantly contrary to prior beliefs, practice is not greatly modified. In fact, the positive nature of feedback from experience may encourage even broader use of the therapy.

There are at least two major problems with these "own experience trials". First are potential biases in measures of association or effect which are introduced due to differential selection of patients into the treatment group. Second and equally important, especially in cases where the occurrence of disease is low, there is a high likelihood that a practitioners decision will be based on an inadequate number of observations.

The biases potentially arising from the use of uncontrolled, unrandomized study designs have been widely recognized and discussed in the literature. The representativeness of the experimental treatment group is essential if one wishes to generalize the results of specific patient cases to all patients. Randomization has been shown to be an effective method for reducing the chance that confounding characteristics of the patient group will bias conclusions. A practice thought to be effective in an individual practice setting may well be found otherwise when subjected to more rigorously designed clinical trial research.

The second important issue addressed in the design of clinical experiments is the determination of an adequate sample size. As was shown by Freiman, Chalmers, et al., insufficient numbers of clinical observations can lead to the improper rejection of superior forms of therapy.¹⁶ The confidence with which any statement can be made increases with the amount of data there is to back it up, and there are quantitative rules for assessing the adequacy of the sample size. These formulas are based on the underlying incidence rate of the phenomenon

being studied, the expected observable change due to intervention, and the decision-makers required level of confidence.

Temin's work on physician drug prescribing behavior casts serious doubt on the value of "learning by doing" for at least one type of medical practice.^{11,17} Using data from a national prescription survey, he found that doctors use each different drug product (different drug or different brand of drug) an average of less than 12 times a year or approximately once a month. He also observed that a wide variety of drugs were used to treat similar diseases. The infrequency with which doctors use each drug product means that they cannot expect to learn about the relative merits of competing drugs and competing brands of the same drug by using them. Even if they have complete enough records to separate the patients to whom each different drug or brand had been given, it would take a very long time to accumulate a large enough sample to show significant differences. In addition, if the drugs or brands to be compared were not prescribed randomly, the effect or results of even a large sample would be impossible to interpret.

In the same vein, Finkelstein and colleagues considered whether surgical practitioners were basing their evaluation of efficacy of preventive measures for pulmonary embolism on an adequate number of observations.¹⁴ Another way of asking the question is whether the low incidence of post-operative thromboembolism make it likely that an individual surgical practitioner would observe enough cases to make a statistically sound decision. Based upon data derived from a national random sample of surgical practitioners in which they were asked to estimate the incidence of pulmonary embolism following surgery, the fraction of surgical patients receiving prophylactic measures against

pulmonary embolism, and the number of patients seen per surgical practice per year, it would take nearly 7 years for a typical practitioner to "validate" the efficacy of one of those measures through "learning by doing."

D. Mechanism of change in clinical practice

It seems reasonable to propose a mechanism through which medical evidence leads to changes in clinical practice. One underlying theory starts with the elementary marketing notion that you have to make people notice a new product before you can convince them to buy it. The mechanism, therefore, is two-fold. First, people become aware of the existence of a recommendation, and second, they are persuaded to follow it.⁷

Awareness is an issue because of the multiplicity of information in the world today. Doctors tend to be overwhelmed with professional information of many sorts and cannot possibly absorb all of it. Largely because of this sea of information, physicians act according to what Temin has called the "customary" mode of behavior, which reflects what doctors consider to be the norms of their profession, and which may have been communicated to them through formal (textbooks, journals, seminars) as well as informal (person to person) means.¹⁷

Once people have become aware of a conflict between their behavior and some external stimuli, they can shift out of their customary behavior into "instrumental" behavior in which they reexamine their behavior and seek to bring it into conformance with the external environment.¹⁷

Instrumental behavior is determined by its effects, while customary behavior is shaped by the norms of the behavior itself. For example, a doctor acting instrumentally might avoid performing "experimental"

procedures on patients because of having learned of clinical trial findings disputing their effectiveness. While a doctor acting customarily may automatically include a particular diagnostic test as part of a clinical algorithm for dealing with a presenting problem because of it is part of the normal practice at the hospital or in the group.

To describe the mechanism by which technology assessments might have their impact, therefore, we have to answer three questions. First, how do people acting in the customary mode of behavior change behavior? Second, how do people in the instrumental mode change behavior? And third, when do people switch from one mode of behavior to the other?

The question of behavior change within the customary mode is of interest because technology assessment could have an impact without causing people to shift from customary into instrumental behavior. Customs do change over time in response to shifts in the social structure of the medical community and technology assessment could be the cause of a significant--albeit gradual--effect by this means. Since customs change slowly, however, it is likely that we may not be able to verify directly the impact of technology assessments on behavior purely within the customary mode, particularly of those held in the very recent past.

The question of behavior change within the instrumental mode is important because we need to know how people utilize information once they become aware of its importance. Reports of technology assessments are only one form of medical information. We need to know how they are received relative to other forms, such as advertisements by manufacturers. In addition, Kahneman and Tversky and others suggest that people respond differently to the same information if it is expressed or "framed," in their terminology, differently.^{18,19}

The third question--do people switch modes as a result of technology assessments?--is perhaps the key. It is the area we know least about and the area in which we suspect the importance of technology assessments is determined. Behavior change from customary to instrumental mode may take place rather rapidly. Reports of technology assessments appear among a wide variety of information sources. We need to ask which sources of information command attention. The answer may vary according to the subject matter, the nature of the conclusions, or time. It may vary according to the location of the recipient. And, of course, it varies according to characteristics of individuals.

Temin has shown, for example, that doctors' drug prescribing practices generally are determined in a customary mode.¹⁷ That is, they have strong continuity over time; they change in response to social pressure within the medical community; and they are only indirectly related to effectiveness. Prescribing practices are evaluated by reference to a medical custom or guideline, not by examining the therapeutic result.

Continuing with the drug prescribing example, it is likely that neither the medical literature nor doctors' own experience give doctors the information needed to make an instrumental choice between closely competing drugs or between competing formulations (brands, generics) of a single drug. Doctors therefore rely on customary behavior to make their choices. They respond only to information that comes to them through their social contacts in the medical community. This is the lesson of Coleman, Katz and Menzel's classic study. The new drug, gammanym, was adopted without reference to its comparative effectiveness (relative to its earlier competitors, alphany and betany), but in accordance with

the personal contacts within the medical community. This demonstrates both how change takes place in the customary mode and that change is possible in this mode.

Doctors switch into an instrumental mode when there is an increase in the disharmony between their prescribing habits and the therapeutic context. This could come about either because of a conflict within a doctor's own practice or because of a conflict exposed by the media. One important and open question is how to describe or to design information that will lead doctors to shift from the customary to the instrumental mode and actively examine alternative therapies.

IV. TECHNOLOGY ASSESSMENTS TO NARROW THE GAP?

A number of policies calling for technology assessments have been implemented or proposed whose main objective is to enhance the relationship between scientific evidence and clinical practice. Many people believe that these assessments can also contribute to reducing costs associated with the use of technologies with questionable efficacy. However, the particular forms of technology assessment that we will examine address cost-effectiveness issues only indirectly.

Some forms of technology assessment are binding on the practitioner in that access to the technology or reimbursement for using it either would be granted or denied. Other kinds of assessments are largely voluntary and depend on the development of a consensus within the professional community. Technology assessment policies can also be differentiated based on whether they generate a new information base or re-evaluate an old one; and whether they have an active or passive dissemination component.

A. Approval Mechanisms for Drugs and Medical Devices

The Food and Drug Administration approval mechanisms for drugs and medical devices are examples of technology assessments whose results are binding, at least in part, on the manufacturers and on the medical profession. The timing of an assessment is well-defined; it must precede the acceptance of the practice by physicians. The form of the assessment is a clinical trial.

The Federal Food and Drug and Cosmetic Act of 1938 required the FDA to reject an application to market any new drug that the FDA deemed unsafe. The 1962 Drug Amendments to that law transformed the FDA from a passive to an active participant in the treatment of new drug applications (NDAs) -- the FDA had to make an explicit decision on all NDAs, not just intervene with unsatisfactory ones -- and added efficacy to the preexisting requirement for safety for new drugs. The 1962 law also stated more explicitly than the original act how the FDA was to decide whether a drug was safe and effective. But it was not until 1970 that the FDA issued a regulation interpreting the statute to mean a clinical trial.

The 1970 regulation said that the "adequate and well-controlled investigations" cited by the 1962 statute as "substantial evidence" of safety and efficacy had to include a formal test with explicit objectives, selection procedures for both subjects and control groups, observation and recording methods, and statistical analysis. Every study needed a carefully defined control group fitting into one of four possible categories. The drug being tested could be compared with a placebo or another drug known to be active from past studies. Or, the effects of the drug being tested could be compared with the known results

of no treatment or with historical data on the course of the ailment without this drug. The FDA clearly tried to spread its net wide to allow a variety of tests -- so wide in fact that it would be hard to disallow any clinical investigation on the basis of inadequate controls, since the study could be allowed in as an example of historical or no-treatment controls. Clinical experience was ruled out for its informality and lack of explicit objectives and methods rather than its lack of control. Following the 1970 regulations, "substantial evidence" of effectiveness took on a new and expanded meaning. The once good standing of adequate, documented clinical experience was swept away.¹⁷

In 1976, the jurisdiction of the FDA in the assessment of medical technology was extended even further. In that year Congress enacted the medical device amendments which gave the FDA the authority to grant or deny premarket approval for equipment-embodied technologies having human application. For practical purposes, following these laws, nearly any piece of medical technology which used a manufactured product that was not a drug, could now be considered a device. The process for gaining premarket clearance for a medical device was similar to, but on the whole less resource intensive than the process required to gain clearance to market pharmaceuticals. Three classes of devices were established to reflect different degrees of risk associated with clinical applications. All devices currently on the market and all new devices were to be classified in one of these categories. Only the most critical class of devices was to require premarket approval. Manufacturers desiring to market new devices were to consider their prospective products as requiring premarket approval unless it could be shown to have been substantially equivalent to another device already on the market. Since

the enactment of the medical device laws, the FDA has experienced considerable difficulty in classifying devices already in use.²⁰ Many aspects of the process of implementation of these laws still remain uncertain. However, only class three medical devices, before gaining their premarket approval, need to submit clinical trial data such as that required for drugs.

The behavior shown by the FDA is consistent with a desire to play it safe and to err on the side of being, if anything, too conservative, rather than too permissive. For practical purposes, a denial of premarket approval by the FDA means that the product or practice is given no chance to be used and never gets tried by the profession.

The issue of whether FDA approval mechanisms for drugs and medical devices advance the objectives of the policy makers who promulgated the laws has been addressed in the published literature. Some authors have even attempted to compare the costs and benefits of the policies as they affect society.²¹ The costs are those of preventing or delaying access by patients to efficacious and needed products. The benefits would include any cost avoidance that came about through diminished use of products without efficacy. This literature falls short of definitively establishing whether there is a net cost or net benefit to these policies, because the uncertainties involved in estimating both costs and benefits are quite large.

Apart from the issues of costs and benefits, there is much that remains unknown about the clinical trial process required by the FDA in its premarket approval decision-making. The information submitted by manufacturers to the FDA is proprietary and not subject to review and discussion by independent experts. How many of the reports of clinical

trials submitted to the FDA describe studies that have been adequately designed and correctly interpreted? How the FDA uses the results of these clinical trials in their decision making is not publicly well known. It would be of particular interest to learn how the FDA's evaluates relative weights of safety and efficacy information and how the outcomes of the approval process depend on these weightings.

In addition to approving a drug for marketing, the FDA approves the label under which a drug is marketed. Since the label indicates the specific conditions for which a drug is to be used, the FDA is in theory controlling not only which drugs are used, but how they are used. But while there are compelling reasons for doctors to follow the FDA's instructions -- not the least of which is protection from malpractice suits -- this part of the regulation seems to be honored largely in the breach. A study in 1972 found that well over half of the surveyed hospital-based use of cephalexin, allopurinol and propranolol was not in accordance with the drugs' labeling.²² And a recent study of cimetidine usage found that only ten percent of its administrations were in accord with the label.²³ It would be of great value to know if these few examples are typical of physician behavior or exceptions to a normally conscientious observance of FDA approval labels. If they represent the rule rather than the exception, it may be time for the FDA to rethink its procedures with respect to drug labels.

Many believe that it is a foregone conclusion that the very existence of the FDA premarket activities enhances the relationship between medical evidence and clinical practice. However, because much of this process is not open to public scrutiny and because some of it may not be affecting physician behavior, the uncertainties as to whether this is the case need to be recognized.

B. Third Party Reimbursement Coverage Decision-Making^{24,25,26}

The decision-making process by which medical practitioners are either granted or denied reimbursement "coverage" for using new technologies is a second example of a binding technology assessment of the "consensus" variety. This decision is distinct from the "reimbursement" decision in which carriers determine the amount of reimbursement for a procedure or practice. A number of insurance carriers have established systematic structures, usually committees to assist in their coverage decision-making. One of the best established is that of the Medicare program which maintains a "coverage committee" composed of all the physicians on the staff of the Health Care Financing Administration as well as some representation from outside the agency. The committee is chaired by a physician in the senior management of the agency who considers this function to be one of his principal interests and responsibilities.

When the Medicare program receives a request to cover a previously experimental technology, it first initiates a process designed to elucidate the issues involved. Once the coverage staff are certain they understand the questions that need to be addressed, they defer to the Public Health Service via the Office of the Assistant Secretary for Health for an evaluation of the current state of the art and professional consensus regarding its use. The Public Health Service coordinates the "technology assessment by consensus" and serves as a buffer between HCFA whose budgetary interests might best be served by denying coverage and the National Institutes of Health, the agency whose mandate includes the enhancement of the medical applications of research. By statute, a coverage determination can be based only upon the assessment of clinical

efficacy and cannot consider the benefits of using the technology in relation to its costs.²⁵

The Medicare program began to make requests for assessments to the Public Health Service as early as 1975, however, the process did not become routine until near the end of the 1970s. The National Blue Cross/Blue Shield Association, Blue Shield of California and Massachusetts Blue Cross/Blue Shield are among the other organizations that have established coverage committees. By preventing the use of unproven or inefficacious practices, coverage decisions can potentially enhance the relationship between medical evidence and clinical practice, and at the same time have a favorable impact on costs.^{25,26}

The coverage decision-making process is very new and, so far as we are able to determine, has not yet been the object of a systematic evaluation of its impact. Many of the questions that would need to be addressed about this process of evaluation by consensus relate to the content of the consensus letters or statements which the Public Health Service receives in response to its inquiries to various opinion leaders and professional societies regarding the use of the practice. It would be important to determine whether the information these technology evaluators are receiving is substantially better than that which is generally available to the practicing physician. Do the coverage committees have access to unpublished research data? To proprietary data? Do they rely heavily on the clinical experience of their advisors? And are there significant differences in opinion about the use of the technology among the various experts or expert groups whose opinion was sought for the consensus?

These above questions are tantamount to asking whether the information assembled as part of the coverage decision process typically contains records of customary behavior, (that is, what people are actually doing) or instrumental information (that is, genuine attempts to assess the efficacy of the technologies.) Do vendors or other identifiable interests influence the process by attempting to saturate the decision-makers with information supportive of their viewpoint? And to what extent is the information sought for the consensus dependent on or independent of the medical researchers or specialists who are professional advocates for the technology or practice?

The coverage decision-making apparatus that has been implemented so far may be having an impact on other consensus bodies or on other third-party carriers. It would be as important to determine whether a negative coverage decision prevents the use of the practice as it would to examine whether a positive decision affects how effectively one is used. Some new procedures will not be that distinctly different from already established and reimburseable procedures. For example, cardiac catheterization is generally reimburseable, while coronary angioplasty (PTCA), a form of heart catheterization may still be considered experimental. We have evidence that some practitioners have performed PTCA, billed it as catheterization, and probably did so without great deception.

C. Clinical Trials

By clinical trials, we refer specifically to the large scale, costly, high visibility, multi-center, randomized, controlled clinical research activity that often has been supported by NIH, VA, or other public funds. These differ from the trials mandated by FDA regulations

in that they are non-binding and typically are conducted after the medical practice or product has become part of a customary pattern of use. Design constraints permit these clinical trials to be done only on a narrow range of therapies for which randomization of patients into experimental and controlled groups is considered ethical. Therapies already believed (but not proven) to be either efficacious or inefficacious cannot be ethically prescribed to one group of patients and withheld from another by random assignment.²⁷ Clinical trials of drugs such as oral drugs for diabetes,²⁸ beta blockers,²⁹ and laetrile³⁰ have been aborted by ethical considerations or done without controls.

Clinical trials may be responsible for changing medical practitioner behavior even if they fall short of definitively establishing effectiveness or lack of effectiveness for the technologies under study. They can have high visibility and their results can be reported prominently to the health profession and in public media even if published reports offer insufficient information to critically evaluate the design of the trial.³¹ Clinical trial findings are often reported in academic medical journals, with a broad readership of generalists and specialists likely to be interested in the information. These official reports can generate editorials, letters to the editor, and secondary articles in the same and other publications. This information may appear later in controlled-circulation journals, which are generally supplied without charge to practitioners, and which may have even larger circulation than the journal in which the primary report was published. Reports of clinical trial findings are also disseminated by pharmaceutical detail personnel who must remain up-to-date about research developments relating to their products. Articles sometimes also appear

in newspapers and magazines that reach the public at large. These formal channels of communications may or may not be effective in promulgating clinical trial results.

Finkelstein and colleagues studied whether the publication of results from two large NIH-supported clinical trials of oral agents for the treatment of diabetes and of high cholesterol coincided chronologically with behavior change of practitioners, and how plausible was the argument that the clinical trials were responsible.³² Both trials produced negative results and both were conducted well after the drugs had become accepted as customary practices.

It was found that the use of both classes of drugs began to decline at a time which corresponded chronologically to the publication of results of the major clinical trial. Before publication, the preponderance of journal articles about these drugs expressed a favorable opinion toward treatment of diabetes or high cholesterol with oral drugs. After the results were published, most articles expressed negative opinions about such treatments.³² The finding that the clinical trial of oral agents for diabetes has impacted on clinical practice appears to contradict reports within the first few years following the completion of the trial which reported that utilization of these drugs was still increasing or at least not declining.³³ Such observations were correct at that time, but we can now observe that the decline in the use of the oral agents for diabetes control in response to medical evidence from the clinical trial seems to have been not only plausible, but rather dramatic.

A recently completed survey conducted for the National Heart, Lung, and Blood Institute has corroborated the observation that physicians were

influenced by clinical trial findings to abandon cholesterol-lowering drugs.³⁴ However, the same NIH survey was unable to offer evidence that the negative findings of a clinical trial of a different practice (aspirin therapy to prevent myocardial infarction) caused a change in clinical practices.

More attention needs to be devoted to understanding the impact clinical trials may be having on physician practice and behavior. The acceptance or rejection of clinical trial findings by practitioners may relate in an important fashion to the nature of practices and products selected to be object of clinical trials. Why are clinical trials proposed for some practices and not for others? Levy and Sondik detail decision making criteria that had been used by the National Heart, Lung and Blood Institute for the initiation of large scale clinical research.³⁵ Important inputs to the decision included the scientific basis of the practice whose efficacy is to be examined, the feasibility of designing the study, and the impact of the underlying disorder on society. It is largely unknown whether other agencies supporting this kind of work or conducting clinical trials in other fields use the same or entirely different philosophies and priorities in their own decision-making.

The impact of clinical trials on practice and behavior easily could differ depending upon whether the findings were consistent with the need to adopt further or to abandon an established practice. It might be argued that effecting abandonment of a practice might prove difficult because of the extent of customary patterns of usage which are in force, although the study of Finkelstein et al does not support this hypothesis.³² Adoption of positive trial findings would be consistent with the direction of the technological imperative and also with the

interests of manufacturers and sales personnel to promote the usage of products. However, differential impact on practice of positive and negative trial results need to be systematically studied.

If a practice change takes place as a result of a clinical trial, it is of interest to determine its character and the mechanism by which it takes place. For example, is it the kind of change that takes place relatively slowly as the clinical findings modify or replace the customary patterns of usage? Or, alternatively, does a clinical trial have the effect of shaking practitioners out of the customary mode of behavior into an instrumental mode in which they are more likely to be influenced by the nature and extent of scientific evidence available about the practice?

D. "Consensus Development"

The development of a consensus among a group of experts is another form of non-binding technology assessment. Used by a number of professional groups in the past, one of its latest incarnations is as the NIH Consensus Development Program managed by the Office for Medical Applications of Research (OMAR). Begun in 1977, its objective has been to encourage the effective utilization, dissemination, and transfer of medical technologies. This type of technology assessment has already been used in at least one other country (Sweden) and is believed to be under consideration in some others.

Over thirty NIH consensus conferences to date have focused on particular technologies, some important aspect of whose use is the object of controversy or uncertainty within the medical community. A group of experts is brought together to review relevant research and the state-of-the-art, and to reach a consensus about its value, and criteria

for determining appropriate uses. Once consensus is achieved, a written statement is prepared.

The next step (and the object of the program) is to actively disseminate the consensus statement via the professional and public media and by word of mouth and hopefully, thereby, to influence behavior and practice. Among the implicit assumptions underlying the Consensus Development Program are that: (1) if the topics for consensus conferences are selected with this criterion in mind, experts will usually be able to achieve a consensus on the basis of research done to date (new research per se is not part of the program) and (2) that once a consensus has been reached, practitioners and the public will be influenced to change their behavior, if the consensus has been effectively communicated. There have been only a few instances in which a consensus has not been reached. The impact of the consensus program on behavior is currently the object of two large NIH-Supported Studies which have not yet reported results. Smaller scale studies of the impact of certain specific consensus conferences suggest that a broad segment of the medical community is aware that certain technologies have been evaluated by consensus conferences. However, the nature of the consensus arrived at by the experts is not as widely known.^{36,37}.

Gauging the effect of consensus conferences in enhancing the relationship between medical evidence and clinical practice is especially difficult because these events do not occur in a vacuum. There may be parallel developments regarding the technologies themselves or others that may be viewed as complements or substitutes.

Many of the unanswered questions about the impact of consensus conferences as a means for technology assessment are parallel to those

posed for the impact of clinical trials. The nature of decision-making criteria that ultimately determine when the consensus form of assessment is selected for a particular practice or technology needs to be more systematically understood. The significance of the timing of consensus conferences before or after a practice has been accepted as customary behavior by practitioners is of interest. And, whether consensus conferences and clinical trials are best thought of as alternatives to one another or as complements has both historical and normative implications.

With respect to the mechanism of possible impact of technology assessment by consensus on practitioner behavior, it appears that decision-makers may be relying on the social process of consensus to codify the customary behavior of the professional target group. A consensus conference might be thought of as potentially generating exciting or startling new information about a practice or technology in question. But it needs to be asked just how much information content there really is at a consensus conference? Do consensus conferences carry with them more information than a typical review article on the same subject in a highly visible professional periodical? Is it likely that a consensus statement is taken sufficiently seriously to cause practitioners to pay more attention to the scientific evidence about the practice than they otherwise would or than they typically do compared to the influence of the practices of their peers?

Finally, it would be of interest to consider whether the active dissemination strategy typically chosen for consensus statements is properly directed and assumes the correct model of professional communication which applies to the largest segment of the intended target

audience, the community of practitioners who may not read journals as widely as do academic medical practitioners.

V. CONCLUSION

One of the principal aims of "technology assessment" is to bring clinical practice in line with the scientific evidence regarding the efficacy of drugs, devices, procedures, and tests. The "assessment" seeks to settle a controversy through analysis, evaluation, or consensus. However, "settlement of the scientific controversy", if that happens, does not by itself necessarily imply acceptance of the outputs of the technology assessment process, as measured by changes in practitioner behavior.

We have reviewed various forms of technology assessment used widely in American medicine today. These forms can be classified in many ways, but among the most meaningful is by whether they are of the clinical trial or consensus type, and whether they are done before or after a practice is widespread. This two-fold classification gives rise to four categories of assessments, and we have described one of each type: FDA approval process (clinical trial/before), third-party coverage (consensus/before), clinical trial (after), consensus conference (after).

The effects of all types are uncertain. We know very little about how physicians utilize information in general. We know too little about the actual impact of past technology assessments. And we lack the answers to the myriad specific questions raised above about the alternative forms of technology assessment.

One reasonable kind of reaction to this uncertainty, or rather to the perception that actual medical practice is not as closely in tune

with the best knowledge as we would like, has been to propose other forms of technology assessment or new organizations to perform them. An example is the independent institute for health care evaluation recently proposed by Bunker and colleagues.³⁸

While new organizations and new forms of assessment may be desirable on a variety of grounds, we cannot expect them to automatically advance our understanding of how technology assessment either works or fails to work. If we were to better understand the mechanism for changing clinical practice in response to technology assessments, new forms of assessment could be implemented more meaningfully. To the extent that a new organization were to merely duplicate the existing types of technology assessments described here, its work would be subject to the same questions.

In addition, there would always be other concerns arising from the creation of a new bureaucracy of technology assessment. We have commented on the institutional bias of the FDA. Would a new organization have its own bias?

More thought should be given to the role of drug and device manufacturers in technology assessment. Much of the literature casts them in an adversary role. But while this frequently litigious role is part of the story, it is hardly all. Manufacturers have their own interests in technology assessment. They want to know whether their products work as advertised. They want to use this confirmation in their marketing. And they want to know of possible hazards and liability problems as soon as possible. The relationship of manufacturers to technology assessment as it is practiced and as it should be practiced is complex.

Along with entertaining proposals for new structures to conduct technology assessments of the type we already know how to conduct, we propose more assessments of the assessments themselves. With greater understanding of the ways in which information about technology is disseminated throughout the medical community, we can design new technology assessments or redesign existing ones to enhance the relationship between scientific evidence and clinical practice and, ultimately, to improve the quality of medical care.

References

1. Eddy DM. Clinical policies and the quality of clinical practice. N Engl J Med. 1982; 307:343-7.
2. Fletcher RH and Fletcher SW. Clinical research in general medical journals. N Engl J Med. 1979; 301:180-3.
3. Fineberg HV. Gastric freezing - a study of diffusion of a medical innovation. Medical Technology and the Health Care System. 1979; Washington, DC: National Academy of Sciences.
4. Public Health Service Assessment of Gastric Freezing. PHS Assessment Report Series 1982; 2(). Office of Health Research, Statistics, and Technology. US Department of Health and Human Services.
5. Wennberg JE, Bunker JP and Barnes BA. Need for assessing the outcome of common surgical practices. Ann Rev Pub H on Health. 1980; 1:277-95.
6. Eddy DM. Analysis of a clinical policy: a case study of mammography. Durham, NC: Duke University, 1981. Center for Health Policy Research and Education. Report 18-15.
7. Rogers EM and Shoemaker FF. Communication of Innovation: A Cross-Cultural Approach. New York: Free Press, 1972.
8. Bauer RA and Wortzel LH. Doctor's choice: The physician and his sources of information about drugs. J Marketing Research, 1966; 3:40-47.
9. Coleman, JS, Katz E and Menzel H. Medical Innovation: A Diffusion Study. Indianapolis: Bobbs-Merrill, 1966.
10. Manning PR and Denson TA. How internists learned about cimetidine. Ann Int Med., 1980; 92:690-2.
11. Temin P. Physician prescribing behavior: Is there learning by doing? Economics of Drug Choice, Washington: American Enterprise Institute. 1981; 173-182.
12. Young DA. Communications linking clinical research and clinical practice. Biomedical Innovation (ed. by E.B. Roberts, et al). Cambridge: MIT Press, 1981; 177-199.
13. Allen TJ. Managing the Flow of Technology. Cambridge: MIT Press, 1977.
14. Finkelstein SN, Simon C, Gammill JF. unpublished communication 1980.
15. Luft HS, Bunker JP, Enthoven AC. N Engl J Med. 1979; 301:1364-69.

16. Freiman JA, Chalmers TC, Smith H Jr. et al. The importance of beta, the type II error and sample size in the design and interpretation of the randomized controlled trial: Survey of 71 "negative" trials. N Engl J Med, 1978, 299:690-4.
17. Temin P. Taking Your Medicine: Drug Regulation in the United States. Cambridge: Harvard University Press, 1980.
18. Tversky A and Kahneman D. The framing of decisions and the psychology of choice. Science 1989; 211:453-8.
19. McNeil BJ and Pauker SG, Sox HC Jr. et al. On the elicitation of preferences for alternative therapies. N Engl J Med 1982; 306:1259-62.
20. deCourcy Hinds M, Medical devices can help cure or kill. New York Times, August 8, 1982.
21. Peltzman S. An evaluation of consumer protection legislation. The 1962 drug amendments. J Political Economy 1973; 81:1046-91.
22. Mundy GR et al. Current medical practice and the Food and Drug Administration. JAMA 1974; 229:1744-48.
23. Cocco AE and Cocco DV. A survey of cimetidine prescribing. E Engl J Med 1981; 304:1281-4.
24. Schaeffer LD. Role of the HCPA in the regulation of new medical technologies. McNeil BJ and Cravalho EG (eds.) Critical Issues in Medical Technology. Boston. Auburn House Press. 1981.
25. Greenberg B and Devzon RA. Determining health insurance coverage of technology: problems and options. Med Care 1981; 19:967-78.
26. Hellinger FJ. Controlling costs by adjusting payment for medical technologies. Inquiry 1982; 19:34-43.
27. Schaefer A. The ethics of the randomized clinical trial. N Engl J Med. 1982; 307:710-24.
28. University Group Diabetes Program. A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. Diabetes 1970; 19:747ff.
29. Beta Blocker Heart Attack Trial Research Group, A randomized trial of propranolol in patients with acute myocardial infarction. JAMA. 1982; 247:1707-14.
30. Moertel CG, Fleming TR, Rubin J et. al. A clinical trial of amygdalin (laetrile) in the treatment of human cancer. N Engl J Med. 1982; 306:201-6.
31. DerSimonian R., Charette LJ, McPeck B., et al. Reporting on methods in clinical trials. N Engl J Med. 1982; 306:1332-7.

32. Finkelstein SN, Schectman SB, Sondik EJ et al. Clinical trials and established medical practices: two examples. Biomedical Innovation (ed. by EB Roberts et al) Cambridge: MIT Press, 1981; 200-216.
33. Chalmers TC. The impact of clinical trials on the practice of medicine. Mt Siani J Med. 1974; 41:753-9.
34. Market Facts Inc. The impact of two clinical trials on physician knowledge and practice (under contract to National Heart, Lung, and Blood Institute) Arlington VA, March 1982.
35. Levy RI and Sondik EJ. Decision-making in planning large scale comparative studies. Ann NY Acad Sci 1978; 304-441-57.
36. Jacoby I. Office of Medical Applications of Research, NIH, personal communication.
37. An Impact Study of the 1978 Consensus Development Conference on Supportive Therapy in Burn Care, National Institute of General Medicine Sciences, NIH, July 1981.
38. Bunker JP, Fowles J. and Schaffarzick R. Evaluation of medical technology strategies. Proposal for an institute for health care evaluation. N Engl J Med 1982; 306:687-92.